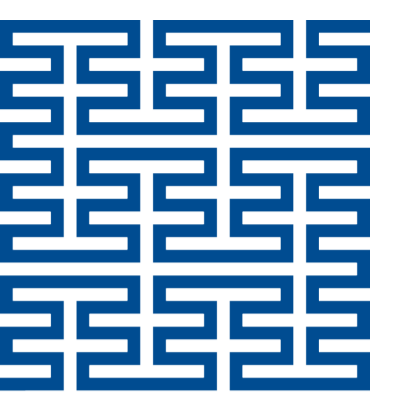


University of
St Andrews

ARCHITECTURE FOR END-TO-END DATA MANAGEMENT

Alan Dearle, Simon Dobson, Shyam Reyal, Simone Conte

(alan.dearle, simon.dobson, smr20, sic2)@st-andrews.ac.uk



School of Computer Science
University of St Andrews

Administration

- Accounting ✓
- Visualization
- Restrictions ✓
- Machines ✓
- Experiments ✓
- Users ✓

Booking

- Authentication ✓
- Submission ✓
- Cancellation ✓
- Mutual Exc. ✓

Machine
Automation ✓

Manual
Upload ✓

Data Archival
and Metadata
Generation

- Security ✓
- Redundancy ✓
- Resilience ✓

Search

- Structural
- Textual ✓

Visualization

- Structure
- Spectrum ✓
- Template

Data Linkage

- Pure. Etc.
- Open Access
- Annotation
Collaboration
Versioning
- Auto. Sync
- Authenticated

Notification

- Email ✓
- Social Media
- Pub. Display

DRAWBACKS OF EXISTING SYSTEMS

- Multi-step process on multiple decoupled systems when booking, running and analysing experiments
- No blanket authentication on multiple systems
- No automatic notification on completion, errors, etc.
- No mechanism to enforce administrative policies
- Data stored in an insecure, non-resilient manner
- No efficient search mechanism - requires browsing
- No remote access – requires physical interaction
- No mechanism for collaboration and data linkage



CASE STUDY - NOMAD

NOMAD – *NMR Online Management And Datastore* – is an active provenance-preserving scientific data store for Nuclear Magnetic Resonance (NMR) Researchers.

We have applied our Data Management Architecture to a real case study, showing what benefits can be achieved using a specific booking and data archival workflow. It was observed that with NOMAD, users ran more experiments, with more efficiency and satisfaction. Administration was also made a lot easier.

6

NMR Spectrometers
In the Lab

34

Research Groups
using NMR

127

Distinct Programmed
Experiments

389+

PhD Students and
Postdocs using NMR

Number of NMR Experiments in 2010/2011 65,846

Number of NMR Experiments in 2011/2012 68,144

Number of NMR Experiments in 2012/2013 (partial integration) 74,742

Number of NMR Experiments in 2013/2014 (full integration) 96,482

>10^{TB}

of Research Data
2003-present

21,387,051

Data Directories

120,473,960

Data Files



2
Admin
Staff

FUTURE WORK

- Integrate with Digital Libraries (i.e. Pure)
- Integrate data from multiple sources (i.e. NMR, Powder X-Ray, Mass-Spec)
- Integrate data across multiple departments (i.e. Chemistry, Physics, Biology)
- Fine-grained accounting and book keeping
- Manual/Automatic research data annotation.
- Enhancing workflow and collaboration
- Structural data search (i.e. SMILE's search for NMR Spectra on Chemical structures)
- Project based experiment archival facilitating reports writing (i.e. Electronic Lab Books)
- Alternative authentication methods (i.e. RFID)
- Integration with Social Media